

AN: PAT 1995-337074  
TI: Recognition of persons from image sequence motion information by processing of typical or predetermined series of movements recorded as digital images for feature vector extension  
PN: WO9525316-A1  
PD: 21.09.1995  
AB: The images of the person exposed to a high-speed camera (10) are converted into a digital sequence from which a flux vector field or sequence is derived and averaged. A sequence of lower resolution is subjected to a three-dimensional fast Fourier transform if the position invariance of the image or person is desired. The result is a one-dimensional feature vector from which a recognition signal (EB) is obtd. by classification with recourse to a learning algorithm. Complementary audio and static image information may be evaluated to provide recognition.;  
PA: (FRAU ) FRAUNHOFER GES ANGEWANDTE FORSCHUNG EV;  
(FRAU ) FRAUNHOFER GES FOERDERUNG ANGEWANDTEN;  
IN: BAUER N; BOEBEL F; WAGNER T; BOEBEL F G;  
FA: WO9525316-A1 21.09.1995; US6101264-A 08.08.2000;  
DE4413788-C1 12.10.1995; EP758471-A1 19.02.1997;  
EP758471-B1 28.07.1999; DE59506480-G 02.09.1999;  
CO: AT; BE; CH; DE; DK; EP; ES; FR; GB; GR; IE; IT; LI; LU; MC;  
NL; PT; SE; US; WO;  
DN: US;  
DR: AT; BE; CH; DE; DK; ES; FR; GB; GR; IE; IT; LU; MC; NL; PT;  
SE; LI;  
IC: A61B-005/117; G06K-009/00; G06K-009/62; G07C-009/00;  
MC: S05-D01C5A; T01-J04B; T04-D07D5; T04-D09; T05-D01B;  
W04-V04A;  
DC: P31; S05; T01; T04; T05; W04;  
FN: 1995337074.gif  
PR: DE4408792 15.03.1994; DE4413788 20.04.1994;  
FP: 21.09.1995  
UP: 21.08.2000

---

Best Available Copy

[10]



⑯ BUNDESREPUBLIK  
DEUTSCHLAND



DEUTSCHES  
PATENTAMT

⑯ **Patentschrift**  
⑯ **DE 44 13 788 C 1**

⑯ Int. Cl. 6:  
**G 06 K 9/62**  
G 07 C 9/00

⑯ Aktenzeichen: P 44 13 788.5-53  
⑯ Anmeldetag: 20. 4. 94  
⑯ Offenlegungstag: —  
⑯ Veröffentlichungstag:  
der Patenterteilung: 12. 10. 95

**DE 44 13 788 C 1**

Innerhalb von 3 Monaten nach Veröffentlichung der Erteilung kann Einspruch erhoben werden

⑯ Innere Priorität: ⑯ ⑯ ⑯  
15.03.94 DE 44 08 792.6

⑯ Patentinhaber:  
Fraunhofer-Gesellschaft zur Förderung der  
angewandten Forschung e.V., 80636 München, DE

⑯ Vertreter:  
Leonhard und Kollegen, 80331 München

⑯ Erfinder:  
Wagner, Thomas, 91058 Erlangen, DE; Böbel,  
Friedrich, 91080 Uttenreuth, DE; Bauer, Norbert,  
91058 Erlangen, DE

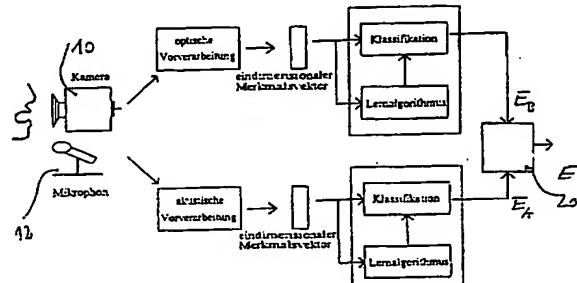
⑯ Für die Beurteilung der Patentfähigkeit  
in Betracht gezogene Druckschriften:

AT 72 083 E  
US 51 64 992  
EP 00 82 304 B1  
EP 05 82 989 A2

Artificial Intelligence, Vol. 17, 1981, pp. 185-203;  
Tagungsband »Artificial Neural Networks and Gen  
Algorithms«, Innsbruck, April 1993, S. 206-212;

⑯ Personenidentifikation mit Bewegungsinformation

⑯ Das technische Gebiet der Erfindung ist die Personenidentifikation. Es wird vorgeschlagen, die Bewegungsinformation aus Bildsequenzen auszuwerten, die sich ergeben, wenn ein typischer oder vorgegebener Bewegungsablauf (2; 2a, 2b, 2c, 2d, 2e) der zu identifizierenden Person ausgeführt wird. Der erwähnte Bewegungsablauf wird aufgezeichnet und mittels eines synergetischen Computers ausgewertet, um hohe Sicherheit bei der Erkennung zu erhalten. Ton- und statische Bildinformation können begleitend ausgewertet werden.



**DE 44 13 788 C 1**

Das technische Gebiet der Erfindung ist die Personenidentifikation mittels Auswertung von Informationen, die der Person an sich anhaften und ohne Verwendung von Hilfsmitteln, wie Ausweise, Schlüssel oder Geheimzahlen.

Das sichere Erkennen — und damit das Identifizieren — von Personen ist nach wie vor technisch ungelöst. Derzeitige Ansätze gehen davon aus, daß sich Personen über Hilfsmittel identifizieren. Beispiele dafür sind der oft gefälschte Ausweis, die verlorengegangene Schlüssel oder die allzu häufig vergessene Geheimzahl. Auch Passwörter haben im heutigen Leben nur eine kurze Lebensdauer, da sie entweder zu einfach gestaltet sind oder von unberechtigten Personen leicht in Erfahrung gebracht werden können. Gerade die zuletzt genannte Gefahr tritt häufig dann auf, wenn man eilig eine Zugangsberechtigung benötigt und das eingegebene (vermeintlich richtige) Passwort nicht funktioniert.

Sicherer Identifizierung von Personen heißt aber auch: Schützen eines lokalen Bereiches gegen unberechtigten Zutritt. Beispiele hier sind militärische Anlagen oder gefährdete Bereiche in Kernkraftwerken. Auch der Zutritt zu Informationen, wie Akten oder Geheimdienstinformationen, die in unserem Zeitalter zunehmend Bedeutung erlangen, fordert eine genau abgestimmte Zugangsberechtigung für Personen, abhängig von ihrer Zuverlässigkeit und ihrem Kompetenzbereich.

Der Stand der Technik kennt die Analyse statischer Bilder, so der Gesichtserkennung, der Erkennung von Handabdrücken oder die Analyse von Fingerabdrücken. Aufgrund der starken Varianz im Aussehen der Person als auch der Schwierigkeit, reproduzierbare Aufnahmeverbedingungen zu schaffen, zeigen diese Systeme eine "Ungenaugigkeit", die noch so hoch ist, daß von einer "sicheren Personenerkennung" nicht gesprochen werden kann. Sicher kann eine Personenerkennung nur dann werden, wenn sie mit (nahezu) 100%iger Sicherheit arbeitet.

Scheinbar mit einer Bildsequenz befaßt sich die EP 82 304 B1; die Bildsequenz, die dort Anwendung findet, ist aber nicht mehr als eine Vorstufe, um das geeignete Momentanbild (Einzelbild) zu lokalisieren und in einem üblichen statistischen Vergleichsverfahren mit vorab gespeicherter Information zu vergleichen. Um das dort als "Schlüsselereignis" bezeichnete Momentanbild während des Sprechens eines Schlüsselwortes aufzufinden, werden akustisch gewonnene Energiesignale ausgewertet und bei Überschreiten eines Schwellenwertes das relevante Einzelbild aus der fortwährend aufgezeichneten Sequenz ausgesondert, festgehalten und ausgewertet. Zur Auswertung dient die Verzerrung eines auf den Lippenbereich projizierten Linien- oder Gittermusters zum Zeitpunkt des Schlüsselereignisses. Also liegt dort eine — mit akustischer Signalauswertung kombinierte — Einzelbildauswertung vor.

Die Erfindung schlägt zur Erreichung einer sicheren Personenidentifikation dagegen die technische Lehre des Anspruches 1 vor. Sie wird ergänzt und konkretisiert mit den Lehren der abhängigen Ansprüche 2 bis 7. Unabhängig davon schlägt die Erfindung zur Implementierung der vorgenannten Verfahren einen synergetischen Computer vor (Anspruch 10). Das Lernverfahren zur Erstellung von kontrasterhöhenden Adjungierten des Audio- und des optischen Merkmalsvektors ist Gegenstand des Anspruchs 8. Der Audio-Merkmalsvektor kann über eine FFT aus dem Spektrogramm; der opti-

sche Merkmalsvektor kann ebenfalls über eine FFT aus dem optischen Fluß berechnet werden (Anspruch 9).

Anstelle des statischen Bildes des Standes der Technik verwendet die Erfindung Bildsequenzen (Bildfolgen). Diese Bildfolgen definieren die Bewegungsinformationen, die über die Berechnung eines optischen Flusses (vgl. Fig. 6 oder Fig. 7) ausgewertet wird. Um die Auswertung sicher zu gestalten, wird ein typischer oder ein vorgegebener Bewegungsablauf verwendet, dessen Bildsequenzen ausgewertet und mit Vergleichswerten verglichen werden. Vor einem Vergleich kann eine Datenreduktion mittels eines Klassifizierungs-Verfahrens erfolgen (Anspruch 7), mit der eine Filterung großer Datenmengen möglich ist.

Die Erfindung, wie sie in ihrem Anspruch 1 mit ihren maßgeblichen Merkmalen umrissen ist, wertet die Sequenz von Bildern aus, in der das Einzelbild "untergeht".

Ein typischer Bewegungsablauf kann eine Gang- oder Kopfbewegung sein (Anspruch 3); ein vorgegebener Bewegungsablauf kann die Bewegung eines Gesichtes oder eines Teilbereiches eines Gesichtes (Gesichtszüge) sein, die beim Sprechen eines vorgegebenen Wortes oder Satzes entsteht (Anspruch 2). Die typischen und vorgegebenen Bewegungsabläufe können vorteilhaft kombiniert werden (Anspruch 4). Die Auswertung der Kombination kann parallel oder ineinander verschachtelt verlaufen. Zusätzlich zur Bildsequenz-Information kann die Auswertung von statischer Bild-Information eingesetzt werden (Anspruch 5).

Die Bildsequenz-Information und die Auswertung von statischer Bild-Information kann mit der Auswertung akustischer Information ergänzt werden (Anspruch 6), um die Zuverlässigkeit (weiter) zu erhöhen.

Ein synergetischer Computer zur Realisierung des Erkennungsverfahrens im vorgenannten Sinn ist Gegenstand des Anspruches 10; er ist z. B. in einem Tagungsband der Tagung in Innsbruck (April 1993, Seiten 206 bis 212) näher erläutert; allerdings nicht für den Zweck des Anspruchs 10. Der Titel dieser Druckschrift ist "Using a Synergetic Computer in an Industrial Classification Problem"; Gegenstand der Tagung und Name des Tagungsbandes ist "Artificial Neural Networks and Genetic Algorithms". Insbesondere unter der dortigen Ziffer 2 wird die hinter einem solchen synergetischen Computer stehende Theorie erläutert, bei der einer der speziellen Algorithmen auf die Lösung einer Matrixinversion zurückgeführt werden kann, anstelle der Integration eines Systems von Differentialgleichungen. Vorgeschlagen wird auch, neurale Netzwerke zu verwenden. Das dortige Anwendungsgebiet war die Erkennung von Kfz-Felgen, die einem Gieß- oder Spritzguß-Verfahren entstammen. Dabei verbleiben Reste von Aluminium (oder anderem Metall) an den Felgen, die eine genaue Erfassung erschweren. Hinzu kommt, daß die Position der zu erfassenden Felgen ebenso wie die Orientierung vielfältig und dem System nicht bekannt ist.

Auf einer anderen Konferenz IECON 1992 in San Diego, California, vom 9. bis zum 13. November 1992 ist in einem Vortrag von Yamasaki und Takahashi ein adaptives Tonerfassungs-System vorgeschlagen worden, bei dem visuelle Effekte ergänzend herangezogen wurden, um auch in der (technischen) Erkennungstechnik den (humaneigenen) "Cocktailparty-Effekt" zu realisieren. Dieser Effekt äußert sich dadurch, daß ein Mensch in der Lage ist, mit seinen Sinnen trotz eines erheblichen Hintergrundgeräusches und eines sehr schlechten S/R-Verhältnisses (S/N-Ratio) bestimmte ihn interessierende Information herauszufiltern. Diese

menschliche - Eigenschaft - ist - heute - weder - genau - erforscht noch kann sie in der hohen Qualität technisch reproduziert werden. Ziel des vorgenannten Vortrages war es also, ein Tonsignal von einer Quelle (einem gesprochenen Satz von einer Person) aus einer Vielzahl von unterschiedlichen Umgebungsgeräuschen herauszufiltern und dabei das Signal/Rausch-Verhältnis zu verbessern. Als Anwendungsgebiet nennt der Vortrag die Erfassung von fehlerhaften Maschinenkomponenten (durch Erkennen "seltsamer" Geräusche) oder die Erfassung von menschlichen Stimmen in lärmender Umgebung (als Mensch/Maschine-Interface, bei dem eine Person der Maschine auf akustischem Wege mitteilt, welche weiteren Schritte sie einsetzen soll). Die im Vortrag erwähnte Technik verwendet neben akustischen Informationen optische Informationen und verbindet diese beiden Informationsquellen, so daß eine "Sensor-Fusion" entsteht. Die Verbindung der beiden Informationen erfolgt in zwei Ebenen, einem Ton-Untersystem und einer Visual-Untersystem.

Anders als die vorgenannten Ansätze verwendet die Erfindung maßgeblich eine Bildsequenz und den optischen Fluß zur Erreichung einer sicheren (hochgenauen) Personenidentifikation. Erst zusätzlich können akustische Informationen herangezogen werden oder statische Bilder ausgewertet werden.

Das Verständnis der Erfindung soll an einem Ausführungsbeispiel vertieft werden.

Fig. 1 zeigt ein Beispiel der visuellen und akustischen Erfassung von Informationen, die von einer schematisch dargestellten Person 2 ausgehen.

Fig. 2 veranschaulicht einen vorgegebenen Bewegungsablauf anhand von (schematisch in Folge dargestellten) Gesichtszügen 2a bis 2e über der (senkrecht verlaufenden) Zeitachse t. Die Gesichtszüge werden von einer Hochgeschwindigkeitskamera 10 aufgezeichnet.

Fig. 3 ist eine Darstellung des optischen Auswertepfades, mit dem die Bildsequenzen verarbeitet werden, namentlich über ein Flußvektor-Feld und ein solches mit reduzierter Auflösung hin zu einem hochdimensionalen Merkmalsvektor, der in ein Klassifikationsverfahren eingespeist wird. Das Klassifikationsverfahren kann in einem synergetischen Computer programmiert sein.

Fig. 4 ist ein mit Fig. 3 korrespondierender akustischer Auswertepfad, in dem die akustischen Signale digitalisiert und als Spektrogramm dargestellt werden. Der Weg führt auch hier über einen eindimensionalen Merkmalsvektor, der in ein Klassifikationsverfahren eingespeist wird, das in einem synergetischen Computer programmiert sein kann.

Fig. 5 zeigt die Kombination der beiden Auswerteverfahren, wobei die optische und akustische Vorverarbeitung als Blockschaltbild jeweils diejenigen Auswerteschritte der Fig. 3 bzw. 4 beinhalten, die vor dem eindimensionalen Merkmalsvektor liegen. Die Ausgangssignale der optischen Klassifikation und der akustischen Klassifikation werden in einer Verknüpfung 20 zusammengefaßt und ergeben das Auswertesignal E, das bei Verwendung zur Personen-Identifikation die Aussage ist, ob die das Kennwort sprechende Person eine der vorgespeicherten Personen ist.

Die Fig. 6 veranschaulicht anhand einer Lippenfolge beim Sprechen einer Person den Bewegungsablauf, der zum optischen Fluß von der Schaltung gemäß Fig. 3 umgesetzt wird. Die weißen Richtungspfeile sind charakteristisch für die Bewegung bestimmter Bereiche der Lippengegend.

Fig. 7a, 7b und 7c veranschaulichen einen beispielhaften anderen Bewegungsablauf, hier zwei Bilder einer aufgezeichneten Bildsequenz und das sich daraus ergebende Flußvektor-Feld in Fig. 7c. Deutlich ist die Bewegung des Balles mit Richtung und Rotation erkennbar, ebenso wie die Bewegung des Armes des Kindes, obwohl in den beiden nebeneinander gestellten Momentanbildern gemäß Fig. 7a und 7b nahezu kein Unterschied erkennbar ist. Deutlich — geradezu überdeutlich — wird der Bewegungsablauf also mit dem Flußvektorfeld gemäß Fig. 7c, das mit der Schaltung gemäß Fig. 3 berechnet werden kann.

In der Fig. 1 ist eine Person schematisch mit 2 dargestellt. Ein vorgegebener Bewegungsablauf der Gesichtszüge der (schematisch dargestellten) Person 2 ist in Fig. 2 mit 2a bis 2e veranschaulicht.

Das System zur Erkennung der Person 2 verwendet in diesem Beispiel eine Hochgeschwindigkeits-Kamera 10, die auf die Person 2 gerichtet ist, sowie ein Mikrofon 12, das Sprachsignale aufzeichnen kann. Die optischen Signale der Kamera 10 und die Tonsignale des Mikrofons 12 werden einer Ansteuerelektronik 11 zugeführt. Diese Ansteuerelektronik 11 kann die Signale so aufbereiten, daß sie einem Rechner 1 zur Auswertung zugeführt werden können. Dabei ist es sowohl möglich, der Ansteuerelektronik bereits Analog/Digitalwandler bei zu geben als auch diese Analog/Digitalwandler in der Auswerteeinheit 1 selbst anzurufen, so daß die Kopplungen zwischen den Bildinformations-Quellen 10 und 12 auf analoger Basis vorgenommen werden.

Der Verfahrensablauf im Beispiel der Fig. 1 und 2 beginnt mit der Vorgabe eines Wortes, Satzes oder Kennwortes, z. B. dem Namen der zu erkennenden Person. Die Hochgeschwindigkeits-Kamera 10 und das Mikrofon 12 nehmen die dabei entstehenden Signale gleichzeitig auf und leiten sie der Ansteuerelektronik 11 zu. Beim Sprechen des Wortes oder Satzes werden die Bewegungen des Gesichts oder eines Teilbereiches davon aufgezeichnet. Ein Klassifikationsverfahren ist in der Auswerteeinheit 1 programmiert, um zu großen Datensätzen auf ein auswertbares Maß zu reduzieren. Zuvor oder danach wird der optische Fluß der Gesichtsmimik errechnet und anhand von Vergleichswerten mit hoher Sicherheit eine der "gespeicherten Personen" wiedererkannt. Wird keine Übereinstimmung zwischen den vorgespeicherten Daten und der den Zugang oder Zutritt anfordernden Person erkannt, so besteht die Möglichkeit der Wiederholung der vorgenannten Vorgehensweise. Bei höheren Sicherheitsstufen kann es auch ratsam sein, diese Wiederholung nicht vorzunehmen und die Berechtigung sogleich abzulehnen.

Detaillierter ergibt sich die Auswertung in der Auswerteeinheit 1 dadurch, daß der optische Fluß aus der Bildsequenz der Hochgeschwindigkeitskamera 10 berechnet wird, der die Bewegungsinformation in komprimierter Form — jedoch nicht beliebig genau — wieder gibt. Die Bewegungsinformation setzt sich dabei aus einer Mehrzahl von Abtastwerten der Gesichtszüge zusammen, so daß fiktive Punkte des Gesichts eine bestimmte Bewegung oder Bewegungskurve beschreiben, die mit dem optischen Fluß (vgl. Horn und Schunk, "Determining Optical Flow", Artificial Intelligence, Vol. 17, 1981, Seiten 185 bis 203) festgelegt werden. Die akustischen Daten werden mit einer Kurzzeit-Fourier-Transformation (FFT) ausgewertet, um die vorhandene Frequenzinformation adäquat darzustellen.

Das in der Auswerteeinheit 1 programmierte Verfahren und die Ansteuerelektronik 11 der Fig. 1 und 2 fin-

den sich in den Fig. 3, 4 und 5 wieder. Eingangsgröße in Fig. 3 ist das optische Signal über die Kamera 10, Ausgangsgröße ist eine erkannte Person oder ein Erkennungssignal  $E_B$ . Eingangsgröße in Fig. 4 ist ein akustisches Signal über das Mikrofon 12, Ausgangsgröße ist das Erkennungssignal  $E_A$ . Beide Ausgangsgrößen können in einer Verknüpfung 20 kombiniert werden und bilden das Ausgangssignal  $E$ , das für eine erkannte Person stehen kann. Die Verknüpfung 20 kann als einfaches UND-Gatter ausgestaltet sein.

Das in Fig. 3 dargestellte optische Verarbeitungssystem verarbeitet eine echte Bildsequenz und diese Bildsequenz repräsentiert sich in dem eindimensionalen Merkmalsvektor, der dem Klassifikationsverfahren zugeleitet wird. Die aufeinanderfolgenden Bilder der Kamera 10 werden in eine digitale Bildsequenz umgesetzt, aus ihnen wird ein Flußvektor-Feld (Flußsequenz) errechnet, die anhand der Fig. 6 und 7 noch erläutert wird; die Flußsequenz ist ein Bild von Vektorpfeilen. Diese entstandene Flußsequenz wird umgesetzt in eine solche mit reduzierter Auflösung (Mittelung), um die Datenmenge zu beschränken.

Die Flußsequenz mit reduzierter Auflösung kann nochmals fouriertransformiert werden, wenn eine Lage-Invarianz des Bildes oder der Person erwünscht ist.

Das Ergebnis ist der eindimensionale Merkmalsvektor, der der Klassifikation zugeführt wird und aus dem das Erkennungssignal  $E_B$  abgeleitet ist.

In vergleichbarer Weise arbeitet die akustische Auswertung gemäß Fig. 4. Das digitalisierte Akustiksignal bildet ein Spektrogramm, bei dem die hohen Amplituden hell dargestellt sind und die geringen Amplituden dunkel erscheinen. Jede Linie ist repräsentativ für eine Abtastzeit.

Auch hier kann eine Koordinaten-Transformation erfolgen, für akustische Signale bietet sich eine logarithmische Transformation an. Mittels dieser Transformation erreicht man eine Tonhöhen-Invarianz, die vergleichbar ist mit der Lage-Invarianz bei der optischen Auswertung gemäß Fig. 3.

Der eindimensionale Merkmalsvektor, der für die akustischen Signale gemäß Fig. 4 gebildet wird, wird der Klassifikation überreicht, die daraus das Ausgangssignal  $E_A$  bildet.

Die Klassifikations-Verfahren der Fig. 3 und 4 können auf einem synergetischen Computer programmiert sein.

Die Kombination der Verfahren gemäß Fig. 3 und 4 ist in Fig. 5 dargestellt. Jede Datenquelle (akustisch, optisch) extrahiert auf dem beschriebenen Weg geeignete Merkmale, die die akustische Information bzw. die Bildsequenz-Information bestmöglich repräsentieren. Diese Merkmale werden in den eindimensionalen Zahlenvektoren festgehalten und der Klassifikation zur Bildung der Ausgangssignale  $E_B$  oder  $E_A$  übergeben.

Bevor auf die Auswertung in der Klassifikation und den dort auch programmierten Lernalgorithmus eingegangen wird, soll der optische Fluß anhand der Fig. 6 und 7a, 7b bzw. 7c erläutert werden.

Der optische Fluß in Fig. 6 ist z. B. derjenige eines Lippenbereiches. Fiktive Punkte haben die weiß dargestellten Bewegungen, die hinsichtlich Länge und Richtung ausgewertet werden können. Bei der Bewegung sind im wesentlichen nur Vektoren des entstehenden Flußvektor-Feldes zu erkennen, die in vertikaler Richtung verlaufen. Anders ist der Sachverhalt bei einer Rotationsbewegung, beispielsweise des Balles gemäß Fig. 7a, 7b. Dort ist in dem Flußvektor-Feld der Fig. 7

erkenntbar, daß der Ball rotiert. Die Flußvektoren sind also in der Lage, jede Bewegung von fiktiven Punkten auszuwerten und selbst bei statischen, kaum erkennbaren Unterschieden, deutliche Vektoren aufzuzeigen, die den Bewegungsablauf festlegen und markant beschreiben können.

Ebenfalls erkennbar ist die Bewegung des Armes des Kindes und des Knie des Kindes, das in den Fig. 7a und 7b dargestellt ist. Der Kopf ist nahezu unbeweglich.

In den Fig. 3, 4 und 5 ist ein Lernalgorithmus angegeben. Er dient der Festlegung der bestimmten, dem System bekannten Personen. Bei einem solchen Lernverfahren auf einem synergetischen Computer wird entweder für den optischen oder für den akustischen Informationspfad die Vielzahl der Merkmale, die für jede Klasse berechnet worden ist, zu adjungierten Prototypen umgerechnet. Für jede Klasse – jede Klasse entspricht einer Person – existiert ein adjungierter Prototyp. Diese Adjungierte hat die Eigenschaft, diejenigen Merkmale in der Bewertung stark zu berücksichtigen, die signifikante Unterschiede zwischen einzelnen Klassen (einzelnen Personen) aufzuzeigen, während andere Merkmale, die in allen Klassen nahezu gleich sind, stark unterdrückt werden. Sie ist "kontrastbildend".

Mittels des Lernverfahrens werden die adjungierten Prototypen berechnet, diese adjungierten Prototypen ermöglichen die zielstrebige Klassifikation von noch unbekannten optischen oder akustischen Merkmalsvektoren.

Ist das Lernverfahren einmal ausgeführt und besteht ein Datensatz an Klassen, so können zu erkennende Personen (ein zu klassifizierender Merkmalsvektor oder ein zu klassifizierendes Muster) leicht erkannt werden. Der Merkmalsvektor wird gemäß den Schaltungen von Fig. 3 und Fig. 4 bzw. Fig. 5 erstellt und dann der Klassifikation übergeben, die auf die gelernten Klassen zurückgreifen kann.

Die bereits gelernten adjungierten Prototypen, die die Eigenheiten jeder einzelnen Klasse (jeder einzelnen Person) zum Ausdruck bringen, werden im Klassifikationsmodus dann dem noch unbekannten Merkmalsvektor zugeordnet und als "schon dagewesen" erkannt oder als unbekannt zurückgewiesen.

Eine vorteilhafte Ergänzung dieser Sequenz- und Akustik-Auswertungen bildet die Auswertung von Einzelbildern der aufgezeichneten Bildsequenz. Als geeignete Einzelbilder haben sich das Anfangs- und Endbild eines kurzen Zeitabschnittes erwiesen, in dem die dynamische Erkennung erfolgt; sie können zusätzlich als Informationsquelle für die Festlegung der zu identifizierenden Person herangezogen werden.

Werden alle drei Datenquellen kombiniert, so können diese drei Sensor-Eingangssignale in ein Klassifikations-Verfahren eingegeben werden, das wie zuvor erläutert arbeitet.

Die vorerwähnte Verfahrensweise zur Erkennung von Personen hat in einem Versuchsaufbau ergeben, daß unabhängig von der Umgebungsbedingung (Hintergrund, Skalierung, Lichteinfluß) zuverlässige Identifikationsraten erhalten werden.

Besonders vorteilhaft ist an dem beschriebenen Verfahren, daß die zu identifizierende Person keine Hilfsmittel, wie Schlüssel oder Ausweise verwenden muß. Auch sind keine Markierungen oder Kennzeichnungen an der jeweiligen Person erforderlich. Die der Identifikation zugrundeliegenden Bewegungsabläufe, das Sprechen von vorgegebenen Worten oder das Ausführen typischer Bewegungsabläufe, kann häufig gewechselt

werden, so daß eine nicht zugangsberechtigte Person sich auf die jeweilige Erkennungsprobe nicht vorbereitet oder einstellen kann.

## Patentansprüche

5

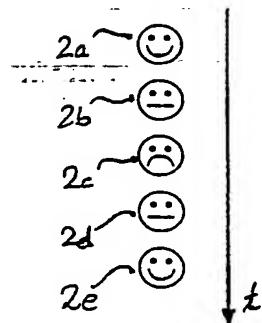
1. Verfahren zur Personenidentifikation, bei dem Bewegungsinformation aus Bildsequenzen (Fig. 2 oder Fig. 6), die aufgrund eines — typischen oder vorgegebenen — Bewegungsablaufs an bzw. von der Person entstehen, aufgezeichnet (10), ausgewertet und mit Vergleichswerten verglichen werden. 10
2. Verfahren nach Anspruch 1, bei dem der vorgegebene Bewegungsablauf die Bewegung der Gesichtszüge der zu erkennenden Person beim Sprechen eines vorgegebenen Wortes oder Satzes ist. 15
3. Verfahren nach Anspruch 1, bei dem der typische Bewegungsablauf eine Gang- oder Kopfbewegung der zu identifizierenden Person ist. 20
4. Verfahren nach Anspruch 2 und 3, bei dem ein typischer Bewegungsablauf mit einem vorgegebenen Bewegungsablauf kombiniert erfaßt wird. 25
5. Verfahren nach einem der erwähnten Ansprüche, bei dem die Auswertung des Bewegungsablaufs (typisch und/oder vorgegeben) mit der Auswertung von statischen Bildern kombiniert wird. 30
6. Verfahren nach einem der erwähnten Ansprüche, bei dem die Auswertung der Bildsequenzen und/oder der statischen Bilder mit der Auswertung von Audioinformation kombiniert (20) wird. 35
7. Verfahren nach einem der erwähnten Ansprüche, bei dem ein Klassifikations-Verfahren vorgesehen ist, daß insbesondere ein Vergleich des unbekannten Merkmalsvektors mit adjungierten Prototypen ist, die aufgrund eines Lernverfahrens erhalten wurden. 40
8. Lernverfahren für ein Personen-Identifikationsverfahren, bei dem
  - (a) zu einem — aus der optischen Bildsequenz (10) und dem aus dem optischen Flußvektor-Feld ermittelten — Merkmalsvektor eine Adjungierte errechnet wird, die als Klasse dem Klassifikations-Verfahren zur Verfügung steht; oder/und 45
  - (b) aus dem akustischen Signal (12) ein Spektrogramm gebildet wird, das in einen Merkmalsvektor umgesetzt und zu einer Adjungierten umgerechnet wird. 50
9. Lernverfahren nach Anspruch 8, bei dem die Merkmalsvektor-Bildung über eine zwischengeschaltete Fourier-Transformation erfolgt. 55
10. Verwendung eines synergischen Computers für eine Personenidentifikation gemäß einem der vorhergehenden Ansprüche. 55

Hierzu 4 Seite(n) Zeichnungen

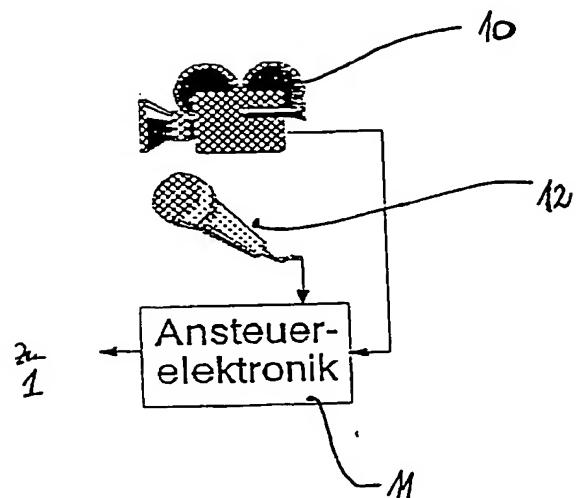
60

65

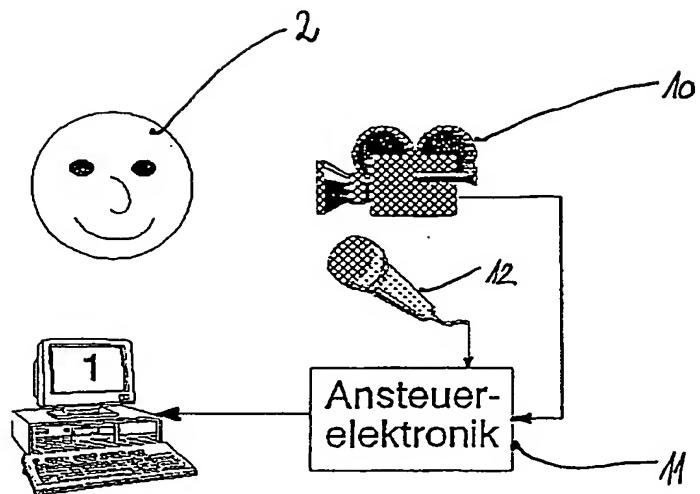
**- Leerseite -**

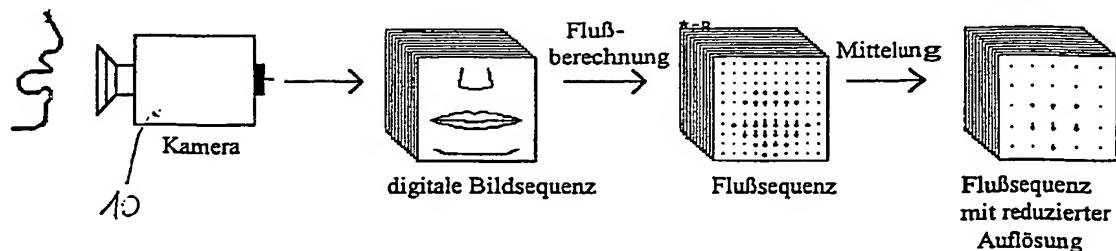


Figur 2

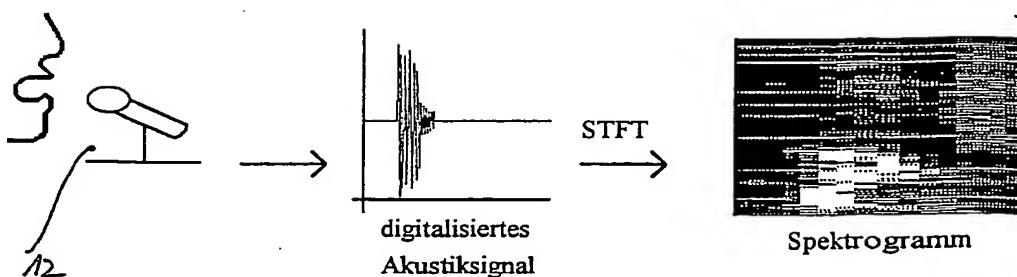
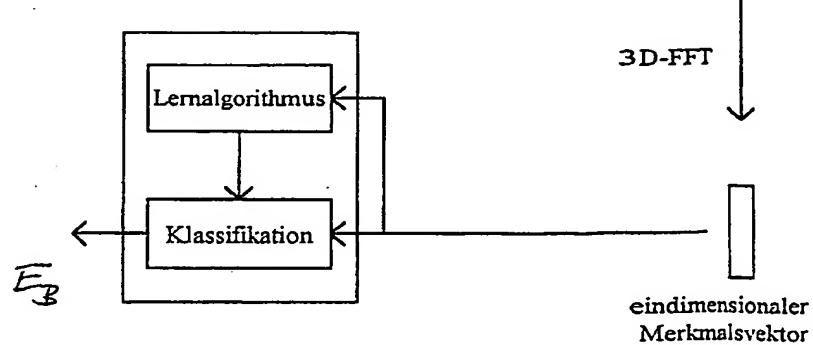


Figur 1

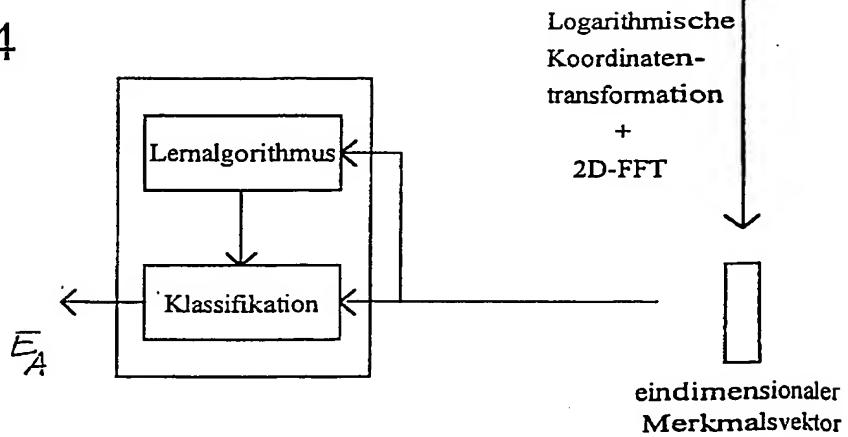


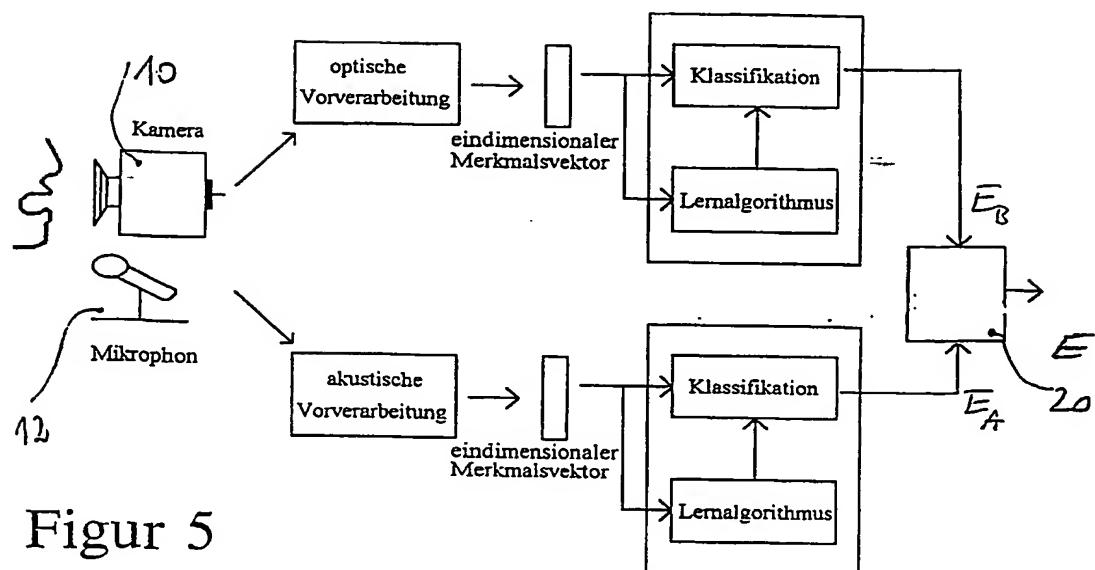


Figur 3



Figur 4

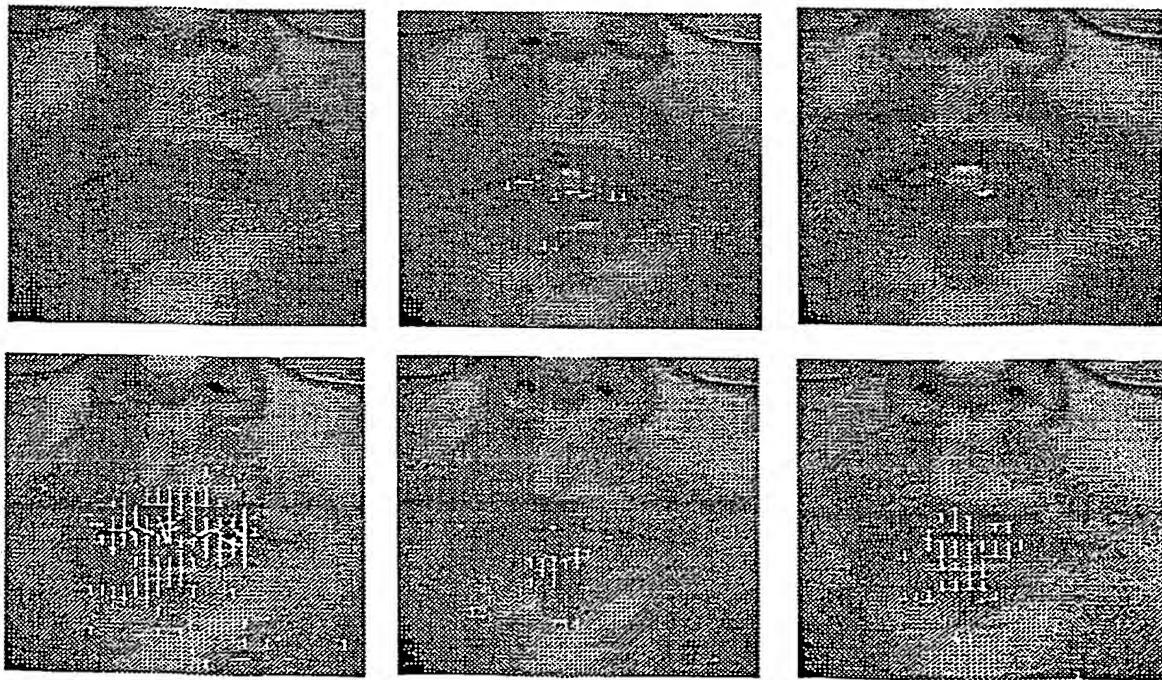


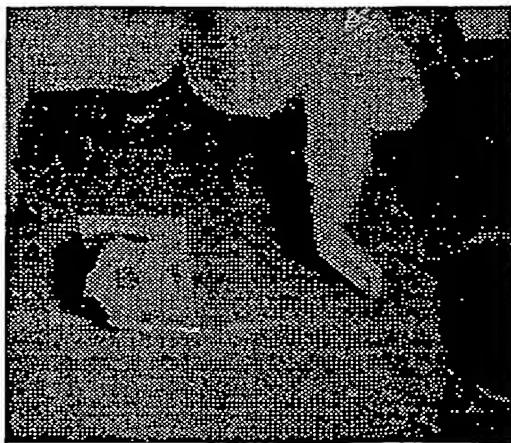


Figur 5

Figur 6

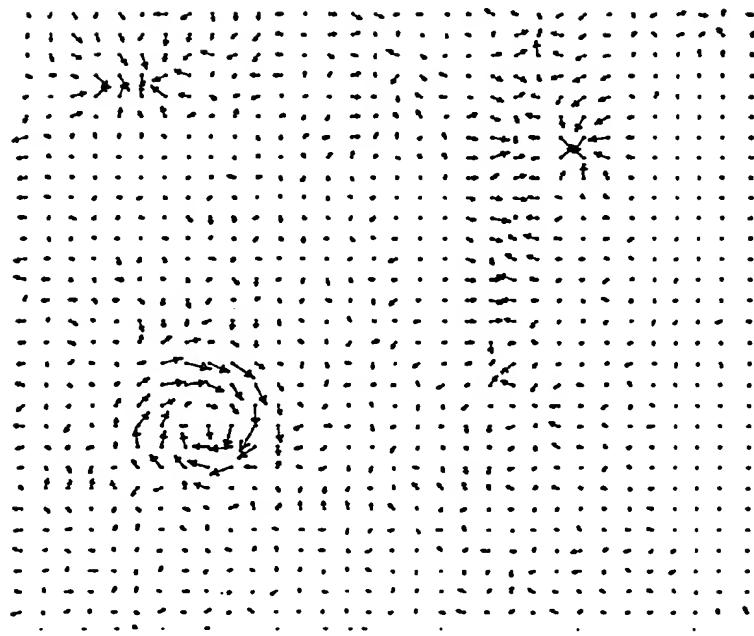
Bewegungsfolge:





Figur 7a

Figur 7b



Figur 7c

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**